

Nicht-reaktive Datenerhebung: Teilnahmeverhalten bei Befragungen mit Paradata evaluieren

Kaczmirek, Lars; Neubarth, Wolfgang

Veröffentlichungsversion / Published Version
Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Kaczmirek, L., & Neubarth, W. (2007). Nicht-reaktive Datenerhebung: Teilnahmeverhalten bei Befragungen mit Paradata evaluieren. In M. Welker, & O. Wenzel (Hrsg.), *Online-Forschung 2007: Grundlagen und Fallstudien* (S. 293-311). Köln: Halem. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-46651-2>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

LARS KACZMIREK / WOLFGANG NEUBARTH

Nicht-reaktive Datenerhebung: Teilnahmeverhalten bei Befragungen mit Paradata evaluieren

Hintergrund

Die computerunterstützte Datenerhebung erleichtert neben der Verarbeitung von Antworten (reaktive Daten) auch die automatische Speicherung von nicht-reaktiven Daten.¹ Im Falle von Online-Befragungen werden dabei beispielsweise Angaben zum Zeitpunkt eines Interviews und dessen Dauer automatisch erfasst. Der vorliegende Beitrag beginnt mit einem Überblick zur bisherigen Diskussion, zu Anwendungsfeldern und dem Nutzen nicht-reaktiver Daten in Umfragen. Der Schwerpunkt wird hierbei auf web-basierte Befragungen gelegt. Eine Typologie schließt den theoretischen Teil ab. Anschließend werden Methoden dargestellt, die es erlauben, Fragestellungen zu untersuchen, die bisher nicht zugänglich waren. Die abschließend dargestellte Untersuchung veranschaulicht den Nutzen von Paradata am Beispiel von Matrixfragen.

Am Beginn dieses Beitrags steht die Unterscheidung zwischen reaktiven und nicht-reaktiven Daten. Bei Antworten in Befragungen handelt es sich um reaktive Daten. Eine Fehlerquelle bei Befragungen liegt in der möglichen Reaktivität, d.h. dass die Messung die Ergebnisse beeinflusst (DIEKMANN 2007; SCHNELL/HILL/ESSER 2004). Bei nicht-reaktiven Daten hingegen ist sich die teilnehmende Person nicht unmittelbar bewusst,

¹ Teilnehmer der berichteten Untersuchung stammten aus dem Panel Sozioland der Respondi AG. Wir bedanken uns für die freundliche Unterstützung durch Dr. Otto Hellwig und Tom Wirth. Des weiteren halfen uns Dr. Wolfgang Bandilla und Dr. Martin Welker mit Revisionsvorschlägen für das Manuskript. Codebeispiele stehen zur Verfügung unter: <http://www.kaczmarek.de/ucsp/> und <http://www.neubarth.net/ucsp/>

dass diese Daten für eine Auswertung zur Verfügung stehen. Bei nicht-reaktiven Daten handelt es sich um Verhaltensspuren aus Handlungsprozessen. Ein Beispiel dafür ist die Befragungsdauer eines Interviews. Den Teilnehmern ist nicht direkt bewusst, dass die Befragungsdauer Teil der Erhebung ist und bei der Datenbereinigung verwendet wird. Im Internet entstehen nicht-reaktive Daten bei fast allen Aktivitäten, z.B. durch den Besuch von Webseiten. Die dabei anfallenden Logfiles stehen ebenfalls als Forschungsgegenstand zur Verfügung. Der vorliegende Beitrag beschränkt sich jedoch auf nicht-reaktive Daten im Kontext von Befragungen, sogenannte »Paradaten«. Als Paradaten bezeichnet man solche Werte, die keine explizite Eingabe seitens der Befragten erfordern (vgl. BOSNJAK/TUTEN 2001; COUPER 1998; HEERWEGH 2003). Sie können automatisch erhoben werden und liefern zusätzliche Hinweise zum *Prozess der jeweiligen Befragung*. Paradaten können daher ebenso gut als Prozessdaten bezeichnet werden. Wir definieren Paradaten daher als von Befragten nicht unmittelbar bewusst anfallende Daten, die auf Personenebene zugänglich sind und zu allgemeineren Aussagen abstrahiert werden können.

Davon abzugrenzen sind Metainformationen. Sie liefern Beschreibungen der Daten und des gesamten Befragungsprojekts (ein ausführliches Klassifikationsschema bietet die *Data Documentation Initiative*, RICHARDSON 2005). Bekannte Beispiele für Metainformationen sind das Codebuch und Projektbeschreibungen, aber auch Ergebnisse aus der Ereignisanalyse: Werden besondere politische oder soziale Ereignisse, die während der Feldphase auftreten, dokumentiert, so lassen sich damit u.U. Veränderungen im öffentlichen Meinungsbild zwischen Beginn und Ende der Feldphase erklären. Ohne Metainformationen bleiben die Daten unverständlich. Eine Zusammenfassung der drei Datenarten bietet Tabelle 1.

Eine Ausdifferenzierung des Begriffes »Metadaten« erfolgte mit der hohen Verbreitung web-basierter Befragungen, da damit eine besondere Relevanz für die Online-Forschung einherging. Durch den Modus der selbstadministrierten computerunterstützten Teilnahme können deutlich mehr Prozessdaten erhoben werden als in einem persönlichen Interview. Die Erweiterung des Datenkonzeptes erfolgte 1998 durch Couper, der weitere Beispiele für Paradaten und die Abgrenzung zu Metadaten anführt (2000: 393f.): »As discussed elsewhere (COUPER 1998), one of the benefits of computer-assisted data collection is that a great deal of automated data is generated by the process and these data can be used in turn to evaluate the process. I term these sources of information »paradata«

TABELLE 1

Paradata, Metainformationen und Daten bei Befragungen

Paradata	Metainformationen	Daten
(Automatisch erhobene) Angaben über den einzelnen Befragungsprozess und die daraus resultierenden Aggregate	Angaben zur Beschreibung der Daten	Angaben und Antworten seitens der Befragten
Beispiele: Dauer jeder Befragung, Antwortzeiten je Frage und Teilnehmer, Anzahl Mausklicks auf einer Seite, Reihenfolge der dargebotenen Fragen inkl. »Zurückblättern«, Seite eines Abbruchs (Drop-Out), Bildschirmauflösung	Beispiele: Codebuch, zusätzliche Beschreibungen der Fragen, allgemeine Projektangaben, wie z.B. durchführende Organisation, Feldzeiten, besondere politische und soziale Ereignisse während der Feldphase	Beispiele: selbstberichtete Angaben, Angaben von dritter Seite, Testdaten. Z.B. Einstellungen, Geschlecht, Einkommen, Bildungsgrad, Persönlichkeitsfaktoren, Intelligenz

(auxiliary data describing the process) to distinguish them from meta-data (describing the data). Sources of paradata include case management information such as response rates, number of calls per case, average interview length, and so on. A most useful type of paradata for usability evaluation is keystroke or trace files (also called audit trails).« Im englischsprachigen Raum ist die Unterscheidung zwischen *metadata* und *paradata* mittlerweile geläufig (vgl. HEERWEGH 2003). Im vorliegenden Beitrag wird die Bezeichnung *Metainformationen* im Gegensatz zu *Metadaten* bevorzugt, um deutlich zu machen, dass solche Angaben nicht im Datensatz erscheinen. Derzeit werden die Begriffe »Para«- und »Metadaten« teilweise synonym genutzt. Nach der folgenden Darstellung möglicher Anwendungsfelder wird zur Klärung der Begriffe das Konzept der Paradata ausführlicher erläutert. Es bleibt zu hoffen, dass sich die Verwendung der Begriffe in Zukunft stärker differenziert.

Anwendungsfelder und Ansätze zu Paradata

Paradata zu Antwortverweigerern und Abbrechern werden am häufigsten berichtet, weil sie in jeder Online-Befragung anfallen. Sie lassen sich ohne zusätzliche Datenspeicherung aus den erhobenen Daten ableiten. Bosnjak und Tuten (2001) verwendeten diese Art von Paradata, um eine Typologie von Antwortverweigerern zu entwickeln. Sind weitere

Angaben zum Antwortverhalten erwünscht, so liegt es nahe, die Antwortzeiten je Fragebogenseite zu speichern (z.B. FRALEY 2004). Für die Zeitnahme pro Frage sind noch genauere Messungen erforderlich, wenn mehrere Fragen auf einer Seite dargestellt werden. Janetzko (1999) stellt dafür ein Skriptsystem zur Verfügung, das Reaktionszeiten zwischen einzelnen Mausclicks erfasst. Eine umfassendere Sammlung von Skripten (The CSP Project, Client Side Paradata) stellt Heerwegh (2003 und o. J.) zur Verfügung. Mit dem CSP-Skript lassen sich neben Antwortzeiten auch Antwortkorrekturen, Pausen und Scrollbewegungen im Fenster feststellen.

Eine ausführliche inhaltliche Diskussion zur Bedeutung von Antwortzeiten und mögliche Definitionen finden sich bei Draisma und Dijkstra (2004) am Beispiel von telefonischen Befragungen. Die Autoren betonen dabei, dass es unterschiedliche Ursachen für ähnliche Antwortzeiten gibt. So konnten die Autoren feststellen, dass eine längere Antwortzeit mit einer höheren Wahrscheinlichkeit für Antwortfehler einhergeht. Zugleich führt eine schnellere Bearbeitung jedoch ebenfalls zu mehr Fehlern. Die Beantwortung schwierigerer Fragen benötigt mehr Zeit, zudem steigt die Antwortzeit mit der Länge der Frage (YAN/TOURANGEAU im Druck), wohingegen die Stabilität einer Einstellung die Antwortzeit verkürzt. Ebenso können kürzere Antwortzeiten auch auf korrektere Antworten in Wissensfragen hindeuten. Somit ist es schwierig, die Ursachen für gegebene Antwortzeiten isoliert zu betrachten.

Weit größere Mengen von Paradata fallen an, wenn Bedienelemente genutzt werden, die nicht dem Standard entsprechen, sondern z.B. Objekte frei auf dem Bildschirm zu bewegen sind (NEUBARTH, in Vorb.). Anwendungen hierfür sind visuelle Analogskalen (VAS), die häufig in der medizinischen Forschung zum Einsatz kommen (BÄLTER/BÄLTER 2005), oder auch die Messung von Präferenzen mit Rangfolgen, die für die Marktforschung von großer Bedeutung sind (z.B. Conjoint-Analysen, WELKER/WERNER/SCHOLZ 2005). Mögliche Paradata sind unter anderem: die Position des ersten Klicks, Verschiebungen der Objekte und die tatsächliche Endposition (vgl. auch FUNKE/REIPS in diesem Band).

Darüber hinaus lassen sich noch weitere Paradata identifizieren. Haraldsen (2005) verwendet die Anzahl von Fehlermeldungen zur Bildung eines Qualitätsmaßes für Fragebögen. Gemessen wird hierbei, wie viele Fehlermeldungen ein Teilnehmer während der Befragung zu sehen bekommt. Ebenso geht die Anzahl möglicher Fehlermeldungen in die Berechnung ein. Als Fehlermeldungen gelten hierbei alle Auffor-

derungen, die ein Teilnehmer erhalten kann, um eine Eingabe zu überarbeiten (für eine Übersicht zur Echtzeitvalidierung von Eingaben vgl. PEYTCHEV/CRAWFORD 2005). Das können zum Beispiel unsinnige Zahlenangaben sein oder Buchstaben in Feldern, in denen nur Zahlen erlaubt sind. Die Qualität eines Fragebogens ergibt sich dabei durch eins minus dem Verhältnis von aktivierten Fehlermeldungen zur Anzahl möglicher Fehlermeldungen relativiert an der Anzahl der Teilnehmer (Gleichung 1).

GLEICHUNG 1

$$\text{Qualitätsmaß} = 1 - \frac{\sum \left(\frac{\text{aktivierte Fehlermeldungen}}{\text{mögliche Fehlermeldungen}} \right)}{\sum \text{Personen}}$$

Nach Haraldsen (2005) wurde dieses Qualitätsmaß mit großem Erfolg zur Verbesserung der Betriebsbefragungen des Statistikamtes in Norwegen eingesetzt.

Insgesamt lässt sich feststellen, dass sich Paradata in der empirischen Sozialforschung für viele verschiedene Zwecke einsetzen lassen (vgl. auch COUPER/LYBERG 2005). Paradata sollen dabei stets dem Zweck dienen, die Qualität von Befragungen zu erhöhen oder zu messen. Sie dürfen nicht dazu missbraucht werden, die Teilnehmer von Befragungen auszuspiionieren (vgl. JANETZKO/HILDEBRANDT/MEYER 2001). Damit einher geht auch die Forderung, dass Paradata nicht ohne Einverständnis von Teilnehmern erhoben werden sollen. Die Richtlinie für Online-Befragungen, welche von mehreren Verbänden getragen wird, besagt deutlich: »Die Befragten müssen vorab ausdrücklich – möglichst durch Anklicken eines entsprechenden Buttons – um ihre Einwilligung gebeten werden, wenn während der Befragung unbemerktbar zusätzliche Informationen (z.B. Typ des benutzten Browsers) erhoben und eventuell gespeichert werden sollen. Dabei sind die Befragten über den wissenschaftlichen Zweck der unbemerkbaren Erhebung und Speicherung zusätzlicher Informationen zu informieren. Die Befragten müssen während der Befragung jederzeit die Möglichkeit haben, ihre diesbezügliche Einwilligung zu widerrufen. In diesem Fall sind die bereits gespeicherten Informationen sofort zu löschen.« (ADM/ASI/BVM/DGOF 2000: 2).

Zur Konzeption von Parادات

Wie oben dargestellt wurde, fasst Couper (2000: 393f.) die Prozessdaten zu einzelnen Befragungen (z.B. Dauer einer Befragung) und deren Aggregate (z.B. durchschnittliche Befragungsdauer) unter dem Begriff »Parادات« zusammen. Mit Hilfe computerunterstützter Datenerhebung und insbesondere in Online-Befragungen lassen sich die einzelnen Datenpunkte mittlerweile bis hinunter zu einzelnen Aktionen wie Mausklicks und Tastatureingaben verfolgen, aufschlüsseln und speichern. Zur besseren Verständlichkeit der verschiedenen Aspekte von Parادات stellt dieser Beitrag daher eine Ordnungshierarchie mit vier Stufen vor (Tab. 2). Daran wird sichtbar, dass mit Parادات verschiedene Aspekte von Daten angesprochen werden, die sich zur Bearbeitung unterschiedlicher Fragestellungen eignen.

Auf der untersten Stufe befinden sich Parادات erster Ordnung. Sie enthalten Prozessdaten zu den einzelnen Aktionen einer Person inner-

TABELLE 2

Die Ordnungshierarchie mit den verschiedenen Aspekten von Parادات. Die hinterlegte Dreieck veranschaulicht das Datenaufkommen und Abstraktionsniveau der verschiedenen Ordnungen.

Parادات vierter Ordnung: Aggregation über Variablen und Personen. Beziehen sich auf die gesamte Befragung und beschreiben diese anhand eines einzelnen Wertes. Beispiele: Teilnahmequote in der Ausschöpfungsübersicht (Response-Rate), durchschnittliche Teilnahmezeit für den Fragebogen

Parادات dritter Ordnung: Aggregation über mehrere Variablen oder Personen. Ersteres bezieht sich auf einzelne Personen und beschreibt deren Teilnahme anhand einer Variable (zeitliche Zusammenfassung), letzteres fasst alle Teilnehmer über ein Konstrukt zusammen. Beispiele: Dauer einer Befragung je Person, Tag des Teilnahmebeginns, durchschnittliche Bearbeitungszeit über alle Teilnehmer für eine bestimmte Fragebogenseite

Parادات zweiter Ordnung: Aggregation mehrerer einzelner Aktionen zu einem inhaltlich bedeutsamen Konzept. Sie beziehen sich auf einzelne Personen und unterteilen deren Teilnahme in verschiedene Zeiträume anhand mehrerer Variablen. Beispiele: Betrachtungszeit einzelner Fragebogenseiten, Anzahl Mausklicks je Fragebogenseite, Anzahl Antwortänderungen je Frage

Parادات erster Ordnung: Erfassung einzelner Aktionen oder Ereignisse während der Fragebogenteilnahme. Beziehen sich auf einzelne individuelle Aktionen. Die Anzahl möglicher Werte ist daher vor Untersuchungsbeginn unbestimmt. Sie sind ohne inhaltliche Bedeutung und technischer Natur. Beispiele: Mausklicks, Mausbewegungen, Tastatureingaben, deren Zeitpunkte und Lokalität

halb einer Befragung. Eine Einzelaktion ist zum Beispiel der Zeitpunkt eines Mausklicks und an welche Stelle des Fragebogens geklickt wurde. Die Tatsache, dass jede Maus- und Tastatureingabe erfasst werden kann, führt leicht zu einer aus Analysesicht unbeherrschbaren Menge an Datenpunkten und Variablen. Dies wird dadurch verstärkt, dass die Anzahl der Paradata auf dieser Betrachtungsstufe nicht vor Untersuchungsbeginn festgelegt werden kann.

Es ist daher hilfreich, bereits im Vorfeld von Befragungen festzulegen, welche Konzepte mit diesen zunächst inhaltsleeren Datenpunkten erfasst werden sollen. Die Konzepte sind Paradata zweiter Ordnung in dem Sinne, dass solch ein Konzept mehrere Datenpunkte erster Ordnung in einem definierbaren Zeitrahmen zu einer inhaltlich bedeutsamen Einheit zusammenfasst. Innerhalb der zweiten Ordnung lässt sich die Anzahl erforderlicher Variablen im Vorhinein bestimmen, wodurch die Datenmatrix bereits vor der Feldphase festgelegt werden kann.

Ein häufig implementiertes Konzept zweiter Ordnung ist die Bearbeitungsdauer je Fragebogenseite. Im Folgenden werden dazu drei verschiedene Operationalisierungen beschrieben. Dazu wird für jede Fragebogenseite im Datensatz eine Variable benötigt, die jeweils die vermutete Bearbeitungsdauer einer Fragebogenseite wiedergibt. Sie kann durch diejenige Zeitdauer bestimmt werden, die vom Seitenaufbau beim Teilnehmer bis zu deren Abschicken durch den Teilnehmer vergeht (ein Ereignis und eine Aktion). Häufiger jedoch wird sie als die Zeitdauer berechnet, die sich aus der Differenz vom Ausliefern der Fragebogenseite bis zum Erhalt der Antworten ergibt. Noch unschärfer würde solch eine Messung mit einer Implementation, die als Zeitpunkte nur die Auslieferungszeiten der Fragebogenseiten verwendet und dadurch die Verzögerungen durch die Datenverarbeitung und die Datenübertragung fälschlicherweise dem Konzept der Bearbeitungszeit einzelner Seiten zuschlägt. Das Beispiel des Konzeptes der Bearbeitungszeit zeigt, dass Zeitmessungen mit Paradata auf unterschiedlich sinnvolle Art und Weise operationalisiert werden können. Weitere Schwierigkeiten können sich bei der Definition des Konzeptes selbst ergeben (z.B. Handhabung unplausibel erscheinender Bearbeitungszeiten und Identifizierung von »Kaffeepausen«). Die Definition der Konzepte erfordert ein umsichtiges Vorgehen, da die verschiedenen Konzepte mit unterschiedlichen Störeinflüssen behaftet sind. Sie erlauben jedoch erst die sinnvolle Auswertung von Paradata.

Paradaten dritter Ordnung verdichten die Variablen zweiter Ordnung nochmals, indem sie diese entweder zu einem Wert je Person *oder* über Personen hinweg aggregieren. Sie können somit Beschreibungen bezogen auf ein gesamtes Interview zusammenfassen. Im Falle des obigen Beispiels entspräche dies je nach verwendeter Definition entweder der gesamten Bearbeitungszeit je Person oder der Befragungsdauer inklusive Serverinteraktion je Person. Gleichwohl existieren aber auch Paradaten dritter Ordnung, die bereits ohne Aggregation abrufbereit vorliegen. Dazu gehören beispielsweise bei einem Abbruch die Seite des Abbruchs, die Uhrzeit und das Datum des Befragungsbeginns, die verwendete Bildschirmauflösung, ob JavaScript ausgeführt werden kann, ob der Teilnehmer Cookies erlaubt und viele andere, die über die korrekte Fragebogenbearbeitung und ihre Rahmenbedingungen Auskunft geben (vgl. KACZMIREK/THIELE 2006).

Die höchste Stufe der Abstraktion bilden Paradaten vierter Ordnung. Sie verdichten die gesamte Erhebung in einem Wert. Solch ein Wert kann die durchschnittliche Befragungsdauer über alle befragten Personen oder eine Teilnahmequote aus der Ausschöpfungsübersicht sein. Auf diesem hohen Beschreibungsniveau beginnt die Trennung zwischen den Bezeichnungen ›Paradaten‹ und ›Metainformationen‹ zu verschwimmen. Sowohl Metainformationen als auch Paradaten vierter Ordnung liefern Beschreibungen, die sich auf die gesamte Untersuchung beziehen. Paradaten vierter Ordnung lassen sich jedoch prinzipiell auf Paradaten niedrigerer Ordnung und damit auf Personenebene zurückführen, was bei Metainformationen (zum Beispiel der Studienbeschreibung) nicht der Fall ist.

Abgesehen von ihrer inhaltlichen Bedeutung und der hier entwickelten Ordnungshierarchie lassen sich Paradaten auch in serverseitige und clientseitige Paradaten einteilen, abhängig davon, ob sie auf dem Webserver oder beim Teilnehmer (HEERWEGH 2003) erhoben werden. Serverseitige Paradaten sind die am häufigsten berichteten Angaben, da die gängigsten zu jedem Befragten vorliegen. Dazu gehören zum Beispiel Zeitstempel für ausgelieferte Seiten, Abbrüche und der Teilnehmerstatus. Genügen serverseitige Prozessdaten nicht, weil Informationen über das Geschehen innerhalb einer Fragebogenseite oder auf Seiten des Teilnehmers gesammelt werden sollen, sind clientseitige Instrumente erforderlich. Dazu gehören beispielsweise Informationen darüber, ob gescrollt wurde und exakte Zeitmessungen einzelner Aktionen. Üblicherweise wird dafür die Technik JavaScript genutzt.

Die hier vorgeschlagene Teilung von Paradata in vier verschiedene Ordnungen erleichtert die Konzeption, Auswertung und Bewertung. Sie gibt eine Struktur vor, um zwischen inhaltsleeren, technisch orientierten Datenpunkten und bedeutsamen Paradata zu unterscheiden.

Universelle Erfassung clientseitiger Paradata (UCSP)

Am Beispiel von möglichen Operationalisierungen der Bearbeitungszeit wurde erläutert, dass sich serverseitige Paradata für die Erhebung vieler Paradata erster und zweiter Ordnung nur bedingt eignen. Ihnen fehlt es sowohl an gewünschter Exaktheit als auch an akkurater zeitlicher Auflösung. Für genaue Prozessdaten sind clientseitige Messungen notwendig. Alle bisher vorgestellten clientseitigen Ansätze (vgl. JANETZKO 1999 und HEERWEGH 2003) haben gemeinsam, dass sie den Einbau von JavaScript-Code in alle zu beobachtenden Bedienelemente (Antwortknöpfe) erfordern. Dadurch wird die Implementierung der Paradatenmessung in die einzelnen Bedienelemente eines selbst erstellten Fragebogens aufwendig und fehleranfällig. Der Einbau dieser Erhebungsfunktionen unter Verwendung von Befragungssoftware ist ebenfalls oft nur mit hohem Aufwand oder gar nicht möglich.

Zudem ist die Datenerhebung ausschließlich auf die vorab präparierten Bedienelemente beschränkt. So können zum Beispiel keine Mausklicks erfasst werden, die neben ein Bedienelement oder außerhalb einer Tabellenzelle erfolgen. Diese wichtigen Informationen zur Bedienbarkeit des Fragebogens stehen deshalb auch nicht für Analysen zur Verfügung. Lediglich erfolgreiche Mausklicks, mit denen Objekte getroffen werden, lassen sich mit den bisherigen Ansätzen erfassen.

Im Gegensatz dazu gibt es jedoch auch eine Möglichkeit zur Erfassung von Paradata, die eine vereinfachte Implementierung – auch in Befragungssoftware – und obendrein die Erfassung von Aktionen außerhalb der Bedienelemente erlaubt. Der vorliegende Beitrag beschreibt daher eine allgemeine bzw. universelle Methode zur Erhebung clientseitiger Paradata (UCSP). Im universellen Ansatz wird, nicht wie bisher die Nutzung von Objekten (Bedienelementen), sondern das Auslösen von Ereignissen (Mausklicks) gespeichert. Dies hat die positive Folge, dass nur eine einzige Anweisung zur Erhebung aller Mausklicks gegeben werden muss. So verringert sich die Anzahl der Code-Änderungen deutlich

gegenüber den bereits bekannten Ansätzen, in denen ein Eingriff für jedes Bedienelement erforderlich war. Da der HTML-Fragebogenstandard eindeutige Namen und IDs der Bedienelemente vorsieht, können die einzelnen Objekte später dennoch differenziert betrachtet werden. Zusätzlich lassen sich auch Fehlclicks analysieren.

Im Folgenden wird ein Codebeispiel vorgestellt, welches die grundlegende Funktionalität veranschaulicht. Dieses Beispiel beschränkt sich auf Paradata erster Ordnung, indem es alle Mausclicks im Fragebogenfenster erfasst. Dieses und ein erweitertes Beispiel, das daraus Paradata zweiter Ordnung ableitet (Anzahl gegebener Antworten, Fehlclicks, wiederholte Klicks), relative Zeiten angibt, Tastatureingaben erfasst, sowie kleinere Erweiterungen enthält, steht auf den Webseiten der Autoren zur Verfügung. Das UCSP-Skript besteht aus zwei Teilen. Der allgemeine Teil (Tab. 4) wird in einer Datei gespeichert. Der zweite Teil (Tab. 3) wird im Hauptteil der Befragungsseite untergebracht und bindet den allgemeinen Teil ein. Der Code aus Tabelle 3 dient dazu, die benötigten Funktionen auf jeder Webseite zu starten und enthält die Zielvariable(n) für die Paradata.

Der Aufbau des UCSP-Skriptes wird im Folgenden näher beschrieben: Der Hauptteil (Tab. 3) definiert eine Variable zur Aufnahme der Paradata, ähnlich dem Anlegen inhaltlicher Antwortvariablen. Da es mit Befragungssoftware nicht immer möglich ist, die Variablennamen frei zu definieren, wird der Variable ein Zeiger (`p_mouse`) zugeordnet. Damit ist der Variablenname im Skript leicht anpassbar. Eine Initialisierungsfunktion startet abschließend die Beobachtung. Die Fragebogenseite kann dann wie üblich um die gewünschten inhaltlichen Fragen ergänzt wer-

TABELLE 3

UCSP-Skriptbeispiel, Hauptteil. Code zwischen den Tags `<body>` und `</body>`

```

<!-- Die folgenden Variablen enthalten die Paradata -->
<input name="v_mouse" id="v_mouse" type="hidden">
<script src="ucsp.js" type="text/javascript"></script>
<script type="text/javascript">
// Variablen von oben merken und Beobachtung starten
var p_mouse = eval("document.forms[0].v_mouse");
init_ucsp();
</script>

```

den, oder das Skript kann in einen bestehenden Fragebogen eingebunden werden.

Der allgemeine Teil (Tab. 4) des Skriptes enthält drei Bereiche: Zunächst werden einige intern benötigte Variablen definiert. Dann folgt eine Funktion, die alle auftretenden Mausklicks verarbeitet und in die im Hauptteil definierte Variable speichert. Der letzte Bereich startet die eigentliche Beobachtung. Das Format der Paradata enthält im Falle von Mausklicks den *Namen* des angeklickten Objekts, einen möglichen *Antwortcode* und die *Zeit* in Millisekunden seit Beobachtungsbeginn, jeweils durch einen Doppelpunkt getrennt.

Ein empirisches Anwendungsbeispiel zur Bedienbarkeit von Matrixfragen und deren Vereinfachung

Das folgende Feldexperiment mit verschiedenen Operationalisierungsformen von Matrixfragen zeigt den Nutzen der oben dargestellten Datenerhebungsmethode. Teil der Fragestellung war es, festzustellen, wie zielicher Teilnehmer in Online-Befragungen die Antworten mit der Maus auswählen, bzw. wie häufig Teilnehmer ins Leere klicken. Dies wäre ohne Paradata nicht zu beantworten. In der Untersuchung mussten die Teilnehmer, wie in Standardfragebögen üblich, die Bedienelemente (Radio-Buttons) direkt treffen, um eine Antwort zu geben. Für das Experiment war den Teilnehmern eine Aktivierung über den Fragentext oder die Umgebung (Tabellenzelle) nicht möglich.

Grundsätzlich lässt sich die antwortssensitive Fläche bei Matrixfragen unter Einsatz von JavaScript auch auf die Umgebung eines Radio-Buttons, d.h. die gesamte Tabellenzelle ausweiten (Tab. 5). Des Weiteren lässt sich die antwortssensitive Fläche unter Einsatz des HTML-Elements ›Label‹ ohne Verwendung von JavaScript auf die Beschriftung ausweiten. Im Beispiel von Tabelle 5 würde ein Klick auf ›weiblich‹ unter Einsatz des Label-Tags ebenso zur Auswahl führen, wie ein Klick auf die rechte Tabellenzelle oder den Radio-Button selbst. Die Verwendung korrekt ausgezeichneter Beschriftungen mit dem Element ›Label‹ ist bisher wenig verbreitet, wird aber bereits für barrierefreie Webseiten generell gefordert (HELLBUSCH 2005). Inwieweit Befragte in Tabellenzellen neben die eigentlichen Bedienelemente klicken, ist bisher ungeklärt und Gegenstand der hier vorgestellten Untersuchung.

TABELLE 4

UCSP-Skriptbeispiel, allgemeiner Teil. Der Code wird in die Datei ucsp.js gespeichert

```
// Datei ucsp.js
// definiere globale Variablen
var start;
var startms;
// Ereignisse für Mausklicks
function mousetracker(e)
{
    // Name und Wert von Ereignis holen
    // Je nach Browser erfolgt die Abfrage unterschiedlich
    var ns6 = document.getElementById && !document.all;
    if(ns6) {
        var p_item=e.target.name;
        var p_item_value=e.target.value;
        if (!p_item)
        {
            if (e.target.id)
                var p_item=e.target.id;
        }
    } else {
        var p_item=event.srcElement.name;
        var p_item_value=event.srcElement.value;
        if (!p_item) {
            if (event.srcElement.id)
                var p_item=event.srcElement.id;
        }
    }
    // Zeit des Ereignisses holen
    var datenow = new Date();
    var nowms = datenow.getTime();

    // Paradata erster Ordnung
    var diffabs = nowms - startms;
    p_mouse.value += p_item + ":" + p_item_value + ":" + diffabs + " ";
}
// Initialisierung und Beobachtung beginnen
function init_ucsp()
{
    start = new Date();
    startms = start.getTime();
    document.onmousedown = mousetracker;
}
```

Die Untersuchung war eingebettet in eine große Befragung, die andere methodische Experimente enthielt. Die Randomisierung erfolgte für jedes Experiment unabhängig, um mögliche Carryover-Effekte zu minimieren. Die Teilnehmer stammten aus dem Pool eines selbstrekrutier-

TABELLE 5

Radio-Buttons in Tabellenzellen mit Beschriftung

männlich	<input type="radio"/>
weiblich	<input type="radio"/>

Es existieren drei Operationalisierungen. Im einfachsten Fall muss für eine Antwort der Kreis direkt angeklickt werden. Unter Verwendung des HTML-Elements »Label« kann auch auf die Beschriftung geklickt werden. Mit Hilfe von JavaScript lässt sich der antwort-sensitive Bereich auf eine Tabellenzelle erweitern. Die Tabellenrahmen sind zum besseren Verständnis in diesem Beispiel sichtbar. Ein vollständiges Codebeispiel ist auf den Webseiten der Autoren einsehbar.

ten Panels (Sozioland) der Respondi AG. Insgesamt wurden 4987 Personen eingeladen an einer Online-Befragung zum Thema »Sicherheit im Internet« teilzunehmen. Die Feldzeit dauerte vom 31. Januar bis zum 12. Februar 2006. Insgesamt 54,5 Prozent der Teilnehmer waren weiblich, 45,5% männlich. Das Durchschnittsalter lag bei 25-29 Jahren. Von den 2003 Personen, welche die Befragung starteten, erreichten 1581 Personen das Ende der Befragung (79%).²

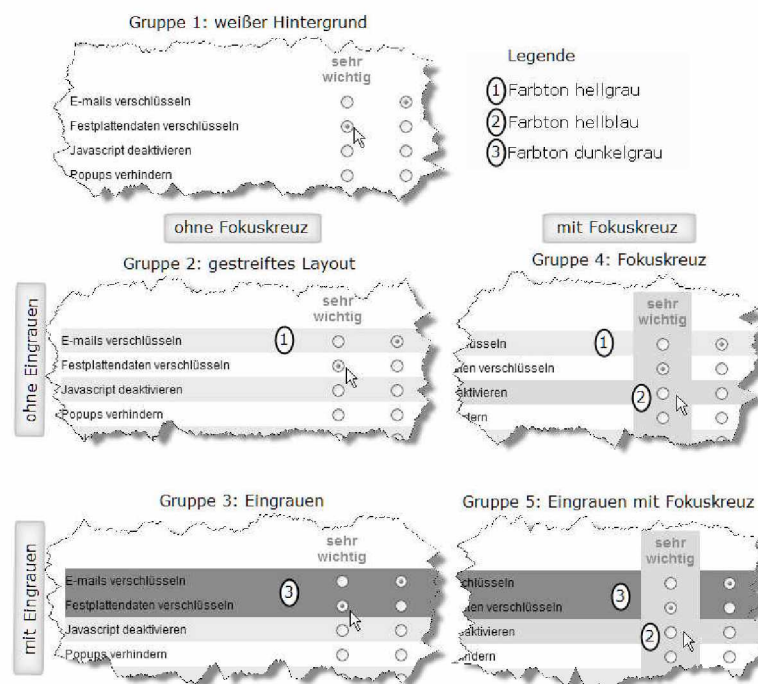
Das experimentelle Design bestand aus fünf Gruppen, die sich im verwendeten Layout und den eingesetzten interaktiven Elementen unterschieden (Abb. 1). Inhaltlich erhielt jede Gruppe die gleichen drei Seiten mit Matrixfragen. Die ersten beiden Matrixfragen dienten dazu, die Teilnehmer mit der jeweiligen Darstellungsform vertraut zu machen. Erst die dritte Matrixfrage wurde ausgewertet. Trainingsphasen sind bei ungewohnten Fragetypen oder -formaten durchaus üblich und führen zu valideren Ergebnissen (GRUNERT 1983). Sie dienen hier der Minimierung von Effekten aufgrund anfänglicher Überraschung und Exploration der möglicherweise unbekannten Umgebung. So zeigte sich im Pretest, dass besonders das Fokuskreuz (Abb. 1, Gruppe 4 und 5) anfänglich zu spielerischen Bewegungen der Maus und inhaltsleerem

2 Die optimale Teilnehmerzahl für ein Experiment ergibt sich nach COHEN (1992) aus der erwarteten Effektstärke, dem Alphaniveau und der erwünschten Teststärke. In der Untersuchung wurde ein kleiner Effekt erwartet, wie dies bei Fragestellungen in den Sozialwissenschaften meistens der Fall ist (COHEN 1988: 284). Um solch einen Effekt zu entdecken (bei einem akzeptablen Betafehler von 0,2 und Alphafehler von 0,05), werden optimalerweise etwa 1200 Teilnehmer benötigt (FAUL/ERDFELDER 1992). Die Zahl der Einladungen ergibt sich damit direkt aus der benötigten Teilnehmerzahl, dividiert durch die erwartete Teilnahmequote. Die tatsächliche Teilnehmerzahl war daher der Fragestellung angemessen.

Klicken einlud, was bei einer Analyse zu irreführenden Schlussfolgerungen führen würde. Als Datengrundlage diente eine Seite mit 16 Items, die jeweils auf einer siebenstufigen Ratingskala plus einer ›Weiß-nicht‹-Kategorie zu bewerten waren.

ABBILDUNG 1

Experimentelles Design mit den fünf Gruppen für die dritte Matrixfrage: »Für wie wichtig halten Sie die folgenden Maßnahmen zur Steigerung der Sicherheit im Internet?«



Die Frage enthielt 16 Items mit einer siebenstufigen Skala, wobei die Endpunkte beschrieben waren mit »sehr wichtig« und »sehr unwichtig«. Die achte Reihe war beschrieben mit »weiß nicht«.

Gruppe 1 erhielt keinerlei optische Unterstützung. Die Matrixfragen werden vor einem weißen Hintergrund dargestellt. Dies entspricht inhaltlich einer möglichen Verschlechterung der gängigen Praxis. Gruppe 2 entspricht der aktuellen Praxis, die Matrix-Items mit einem

gestreiften Layout zu hinterlegen. Die folgenden Gruppen basieren ebenfalls auf diesem Standard, um mögliche Zusatznutzen der Varianten messbar zu machen. In Gruppe 3 werden beantwortete Items mit einem grauen Hintergrund versehen. Dies zeigt, dass die Antwort für das Item bereits gegeben wurde. Personen in Gruppe 4 erhalten ein blaues Fokuskreuz, das dem Mauszeiger folgt und so die genaue Position einer möglichen Antwort in Bezug auf Itembeschreibung und Antwortausprägung betont. Gruppe 5 kombiniert das Eingrauen mit dem Fokuskreuz. Die Gruppen 2-5 entsprechen damit einem vollständig gekreuzten Experiment mit den Faktoren ›Eingrauen‹ (ja/nein) und ›Fokuskreuz‹ (ja/nein).

Neben den eigentlichen Antworten wurden auch alle Mausklicks erfasst, unabhängig davon, ob diese zu einer Antwort führten oder nicht (Paradata erster Ordnung). Zur Klärung der Fragestellung unterzogen wir die Klicks auf Radio-Buttons und die Klicks in Tabellenzellen für die dritte Matrixfrage einer Analyse (Paradata zweiter Ordnung). Die Ergebnisse aus Tabelle 6 zeigen, dass durchschnittlich 38% der antwortenden Personen mindestens einmal auf eine Tabellenzelle klickten. In der hier vorgestellten Untersuchung waren dies Fehlclicks, da damit keine Antwort gegeben wurde. Die Teilnehmer mussten für eine Antwort auf das Bedienelement (Radio-Button) selbst klicken. In den einzelnen Gruppen variierte der Anteil von Personen, die dadurch mehr als notwendig

TABELLE 6

Fehlclicks auf Tabellenzellen. Dargestellt ist die Anzahl an Teilnehmern, die kein- oder mindestens einmal mit der Maus auf eine Tabellenzelle klickten

	Gruppe 1: weißer Hinter- grund	Gruppe 2: gestreifter Hinter- grund	Gruppe 3: Eingrauen	Gruppe 4: Fokus- kreuz	Gruppe 5: Eingrauen/ Fokus- kreuz	Gesamt
Personen ohne Klicks auf Tabellenzellen	202 (65%)	214 (64%)	239 (64%)	170 (54%)	227 (63%)	1052 (62%)
Personen mit einem oder meh- reren Klicks auf Tabellenzellen	111 (35%)	120 (36%)	132 (36%)	147 (46%)	136 (37%)	646 (38%)
Summe	313	334	371	317	363	1698

klickten, zwischen 35% und 46%. Ein Vergleich zwischen den Gruppen zeigt eine höhere Anzahl an Mausklicks auf Tabellenzellen in der Gruppe mit Fokuskreuz gegenüber den anderen Gruppen (Tabelle 6). Die unterschiedlichen Häufigkeiten werden bei einem Chi-Quadrat-Test signifikant ($\chi^2_4=11,9$; $p=0,018$). Das verwendete Fokuskreuz (Abb. 1, Gruppe 4) erweckt möglicherweise den Eindruck, dass sich das Bedienelement auf die gesamte Zelle erstreckt. Gleichzeitig scheint jedoch das Eingrauen in Gruppe 5 dem entgegenzuwirken.

Dem hohen Anteil an Personen, die irrtümlich oder aus Versehen neben das gewünschte Ziel klickten, war die Teilnahme durch das erforderliche erneute Klicken und genaue Positionieren der Maus demnach erschwert. Es ist daher sehr zu empfehlen, auch die zu den Bedienelementen gehörigen Tabellenzellen als antwortensitive Fläche zu implementieren, so wie dies oben beschrieben wurde (Tab. 5). Eine Erweiterung mit JavaScript und Label-Tags ist daher sinnvoll, wobei nicht vergessen werden darf, dass auch den Personen, die kein JavaScript zulassen, eine Teilnahme möglich bleiben muss.

Zusammenfassung

Der Beitrag beginnt mit der Klärung der Begriffe »Daten«, »Metainformationen«, »Metadaten« und »Paradaten« und beschreibt die begriffliche Entwicklung der letzten Jahre. Der Stand der Forschung wird unter besonderer Berücksichtigung von Antwortzeiten und Qualitätsmaßen zusammengefasst. Dabei gehen die Autoren auch auf die technische Komponente bisheriger Umsetzungen clientseitiger Paradatenerhebungen ein. Die verschiedenen Datentypen der Paradaten werden in einer vierstufigen Ordnungshierarchie vertiefend dargestellt und geordnet. Insgesamt lassen sich sowohl einzelne Aktionen auf Item-Ebene als auch daraus abgeleitete Beschreibungen auf Untersuchungsebene unter dem Begriff »Paradaten« subsumieren. Diese Gliederung ermöglicht eine strukturierte Diskussion der verschiedenen Bereiche, Konzepte und Abstraktionsniveaus über die sich dann mit Hilfe von Paradaten empirische Aussagen treffen lassen. Die anschließend vorgestellte Methode zur Erfassung universeller clientseitiger Paradaten (UCSP) bietet einen technisch allgemeinen Ansatz zur Erhebung. Erläuterungen mit Beispielcode veranschaulichen die Erfassung von Aktionen und damit ver-

bundener Fragestellungen, die dem Forscher bisher nicht zugänglich waren. Gleichzeitig vereinfacht der Code die Implementierung in Fragebögen. Den Abschluss bildet die Darstellung einer empirischen Untersuchung zur Bedienbarkeit von Matrixfragen. Klicks, die das eigentliche Bedienelement (Radio-Buttons) verfehlen, lassen sich mit dem UCSF-Skript erfolgreich erfassen. Eine Schlussfolgerung ist, dass solch unbeabsichtigte oder irrtümliche Fehlbedienungen in fehlerfreies Verhalten umgewandelt werden können, wenn auch die das Bedienfeld umgebende Tabellenzelle als Antwort zugelassen wird. In der vorliegenden Untersuchung wäre so mehr als jedem Dritten Befragten durch diese Verbesserung Zusatzaufwand bei der Beantwortung des Fragebogens erspart geblieben. Denn entsprechend der Empfehlungen angepasste Matrixfragen sind in der Bedienung fehlertoleranter, da diese Art von Fehlern in erfolgreiche Bedienung umgewandelt wird.

Eine Befragung wird so vereinfacht, da die Teilnehmer weniger Zeit und Klicks mit der genauen Positionierung des Mauszeigers verbringen müssen. Inwieweit auch hier gilt, dass bessere Bedienbarkeit zu geringeren Abbruchquoten führt, kann Gegenstand einer weiteren Untersuchung sein. Zusammenfassend lässt sich feststellen, dass sich die strukturierte Betrachtung von Paradata und der vorgestellte UCSF-Ansatz zur Qualitätssteigerung und -sicherung von Online-Fragebögen erwiesenermaßen eignet.

Literatur

- ADM; ASI; BVM; DGOF: *Richtlinie für Online-Befragungen*. 2000
- BÄLTER O.; A. K. BÄLTER: Demands on Web Survey Tools for Epidemiological Research. In: *European Journal of Epidemiology*, 20, 2, 2005, S. 137-139
- BOSNJAK, M.; T. L. TUTEN: Classifying Response Behaviors in Web-based Surveys. In: *Journal of Computer-Mediated Communication*, 6, 3, 2001, <http://jcmc.indiana.edu/vol6/issue3/boznjak.html>
- COHEN, J.: *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale [Lawrence Erlbaum Associates] 1988
- COHEN, J.: A Power Primer. In: *Psychological Bulletin*, 112, 1, 1992, S. 155-159
- COUPER, M. P. : *Measuring survey quality in a casic environment*. Paper presented at the Joint Statistical Meetings of the American Statistical Association, Dallas, TX, 1998

- COUPER, M.P.: Usability Evaluation of Computer-Assisted Survey Instruments. In: *Social Science Computer Review*, 18, 2000, S. 384-396
- COUPER, M. P.; L. E. LYBERG: *The use of paradata in survey research*. Paper presented at the 54th Session of the International Statistical Institute, Sydney, Australia, 2005
- DIEKMANN, A.: *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbek b. Hamburg [Rowohlt] 2007
- DRAISMA, S.; W. DIJKSTRA: Response Latency and (Para)Linguistic Expressions as Indicators of Response Error. In: PRESSER, S.; J. M. ROTHGEB; M. P. COUPER; J. T. LESSLER; E. MARTIN, J. MARTIN; E. SINGER (Hrsg.): *Methods for Testing and Evaluating Survey Questionnaires*, New York [Wiley] 2004, S. 131-147
- FAUL, F.; E. ERDFELDER: *GPOWER: a priori, post-hoc, and compromise power analysis for MSDOS (Version 2.0)* [computer programm]. Bonn [Bonn University, Dep. of Psychology] 1992
- FRALEY, C.: *How to Conduct Behavioral Research on the Internet*. New York [Guilford Publications] 2004
- GRUNERT, K. G.: Magnitude-Skalierung. In: *Marketing – Zeitschrift für Forschung und Praxis*, 5, 2, 1983, S. 108-112
- HARALDSEN, G.: *Using Client Side Paradata as Process Quality Indicators in Web Surveys*. Exploratory Workshop Internet Survey Methodology: Towards concerted European research efforts. Dubrovnik, Kroatien, 2005
- HEERWEGH, D.: Explaining Response Latencies and Changing Answers Using Client Side Paradata From a Web Survey. In: *Social Science Computer Review*, 21, 2003, S. 360-373
- HEERWEGH, D.: *The CSP Project Webpage*. Erreichbar unter: <http://perswww.kuleuven.be/~u0034437/public/csp.htm>
- HELLBUSCH, J. E.: *Barrierefreies Webdesign. Praxishandbuch für Webgestaltung und grafische Programmoberflächen*. Heidelberg [dpunkt.verlag] 2005
- JANETZKO, D.: *Statistische Anwendungen im Internet. Daten in Netzumgebungen erheben, auswerten und präsentieren*. München [Addison-Wesley] 1999
- JANETZKO, D.; M. HILDEBRANDT; H. A. MEYER: Zeiterfassung in Online-Fragebögen. In: THEOBALD, A; M. DREYER; T. STARSETZKI (Hrsg.): *Online-Marktforschung*, Wiesbaden [Gabler] 2001, S. 191-212
- KACZMIREK, L.; O. THIELE: *Flash, Javascript or php? Comparing the availability of technical equipment among university applicants*. Poster GOR06, Bielefeld, 2006

- NEUBARTH, W.: Online measurement of moveable objects. In: GOSLING, S.; J. A. JOHNSON (Hrsg.): *Advanced Methods for Behavioral Research on the Internet*. American Psychological Association, in Vorb.
- PEYTCHEV, A.; S. CRAWFORD: A Typology of Real-Time Validations in Web-Based Surveys. In: *Social Science Computer Review*, 23, 2005, S. 235-249
- RICHARDSON, M. A.: *Data Documentation Initiative*. <http://www.ddialliance.org/> (2005)
- SCHNELL, R.; B. P. HILL; E. ESSER: *Methoden der empirischen Sozialforschung*. München [Oldenbourg] 2004
- WELKER, M.; A. WERNER; J. SCHOLZ: *Online-Research. Markt- und Sozialforschung mit dem Internet*. Heidelberg [dpunkt.verlag] 2005
- YAN, T.; R. TOURANGEAU: Fast Times and Easy Questions: The Effects of Age, Experience and Question Complexity on Web Survey Response Times. In: *Applied Cognitive Psychology*, im Druck, DOI: 10.1002/acp.1331